Web Usage Mining for Adaptive and Personalized Websites

Asem Omari and Stefan Conrad
Heinrich-Heine-University Duesseldorf
Institute of Computer Science
Databases and Information Systems
Duesseldorf, Germany
{omari, conrad}@cs.uni-duesseldorf.de

Abstract

The World Wide Web is an important medium for communication, data transaction and retrieving. Data mining is the process of extracting interesting patterns from a set of data sources. Web mining is the application of data mining techniques to extract useful patterns from web data. Web Mining can be divided into three categories, web usage mining, web content mining, and web structure mining. Web usage mining or web log mining is the extraction of interesting patterns from web log server entries. Those patterns are used to study user behavior and interests, facilitate support and services introduced to the website user, improve the structure of the website, and facilitate personalization and adaptive websites. This paper aims to explore various research issues in web usage mining and its application in the field of adaptive, and personalized websites.

1 Introduction

The World Wide Web is an important medium for communication, data transaction and retrieving. Web mining is the use of data mining techniques to extract useful patterns from the web. That extracted patterns are used to improve the structure of websites, improve the availability of the information in the websites and the way that information are introduced to the website user, and to improve data retrieval and the quality of automatic search of information resources available in the web. Web mining can be divided into three major categories:

1.1 Web Usage Mining

Web usage mining describes the usage of web pages. It mines web log records to discover user access patterns of web pages. Usage data are collected from different sources such as web server side data, client side data, and proxy servers. Server side data are collected from the web server of a site that consists of various types of the logs generated by the log server [Pierrakos *et al.*, 2003]. Client side data are collected from the host that is accessing the website by using a remote agents implemented in Java or Javascript [Pierrakos *et al.*, 2003]. That agents are used to collect information directly from the client such as the time the user is accessing or leaving the website, and the user's navigation history. Proxy servers also use access log to record web page requests and responses from the server [Pierrakos *et al.*, 2003].

1.2 Web Content Mining

Web content mining is mining the data that a web page contains. The contents of most of the web pages are texts, graphics, tables, data blocks, and data records. A lot of research has been done to cover different web content mining issues for the purpose of improving the contents of the web pages, improving the way they are introduced to the website user, improving the quality of search results, and extracting interesting web page contents.
The authors in [Lin and Ho, 2002] propose the *InfoDiscoverer* system to discover informative contents from a set of web pages of a website according to HTML tag TABLE in a web page. The system partitions the web page blocks into either informative, or redundant. Informative content blocks are distinguished parts of the page, whereas redundant content blocks are common parts. This approach yields to the increase of the retrieval and extraction precision, and reduces the indexing size and extraction complexity.
A number of methods to help user find various types of unexpected information from his/her competitors' websites are proposed in [Liu *et al.*, 2001]. The work in [Morinaga *et al.*, 2002] presents a new framework for mining product reputations on the internet. It automatically collects people's opinions about target products from web pages, and it uses four different types of text mining techniques to obtain the reputation of those products. The research in [Davison, 2002] examines the accuracy of predicting a user's next action based on the analysis of the content of the pages requested recently by the user. Predictions are made using the similarity of a model of the user's interest to the text in and around the hypertext anchors of recently requested web pages. The authors in [Liu *et al.*, 2003] propose an algorithm called *MDR* (Mining Data Records in web pages) to mine contiguous and non-contiguous data records. It finds all records formed by table and form related tags, i.e., table, form, td, tr, etc. Such data records are important because they often present the essential information of their host pages.

1.3 Web Structure Mining

Links pointing to a document indicate the popularity of the document, whereas links coming out of a document indicate the richness or the variety of topics covered in the document. Web structure mining describes the organization of the content of the web where *structure* is defined by *hyperlinks between pages and HTML formatting commands within a page* [Cohen, 2003].
Understanding the relationship between contents and the structure of the website is useful to keep an overview about

websites. The authors in [Gedov *et al.*, 2004] describe an approach that allows the comparison of web page contents with the information implicitly defined by the structure of the website. In this way, it can be indicated whether a page fits in the content of its link structure, and identify topics which span over several connected web pages. Thus supporting web designers by comparing their intentions with the actual structure and content of the web page.
Other studies deal with the web page as a collection of blocks or segments. The authors in [Cai *et al.*, 2004] use an algorithm to partition the web page into blocks, by extracting the page-to-block, block-to-page relationship from link structure and page layout analysis, a semantic graph can be constructed over the World Wide Web such that each node exactly represents a single semantic topic, this graph can better describe the semantic structure of the web. The authors in [Cohen, 2003] present a survey of some of the ways in which structure within a web page can be used to help machines understand pages.
In this paper, we introduce an overview of various research issues in web usage mining and its application in the field of adaptive and personalized websites. We give an overview about web usage mining in section 2. In section 3, we discuss several data mining techniques used in web usage mining. Before those data mining techniques are applied to web log server data, several preprocessing steps should be done in order to make web log file data ready to be mined, we introduce those preprocessing steps in section 4. Then, in section 5, we discuss adaptive websites. In order to make websites more effective to website users, they should reflect their interests, knowledge, needs, and goals. This can be done through personalization which is the subject of section 6. In this section, we talk about the use of web usage mining techniques for web personalization. In section 7, we list well-known web usage mining and analysis tools. We conclude this paper, and introduce some research directions in section 8.

2 What is Web Usage Mining

Web usage mining or web log mining is the process of applying data mining techniques to web log data in order to extract useful information from user access patterns. Web usage mining tries to make sense of the data generated by the web user's sessions or behaviors [Kosala and Blockeel, 2000]. The web usage data includes data from web server access log, proxy server logs, browser logs, user profiles, registration data, cookies, and user queries [Kosala and Blockeel, 2000]. Web usage mining tries to predict user behavior while user interacts with the web and learns user navigation patterns. The learned knowledge could then be used for different applications such as website personalization, business intelligence, usage characterization and adaptive websites. The authors [Cooley, 2003] show that web usage mining is not only enhanced by web content and structure but it can't be completed without them. There are two common used approaches for web usage mining process [Borges and Levene, 1999]:

- Mapping the log data into relational tables before an adopted data mining techniques is performed.
- Using the log data directly by utilizing special preprocessing techniques.

Web usage mining process consists of three phases: data preprocessing, pattern discovery, and pattern analysis. Pattern discovery is that set of methods, algorithms, and techniques used to extract patterns from web log file. Several techniques are used for pattern discovery such as statistical analysis, clustering, classification, and sequential pattern mining (see section 3). After patterns are discovered they need to be analyzed in order to determine interesting and important patterns, besides the removal of redundant patterns. Pattern analysis has several different forms such as knowledge query mechanism, visualization techniques, and loading usage data into a data cube in order to perform Online Analytical Processing OLAP operations [Srivastava *et al.*, 2000].

3 Web Usage Mining Techniques

In this section, we discuss data mining techniques that are mostly used in web usage mining such as statistical analysis techniques, clustering, classification, association rule mining, and sequential pattern mining.

3.1 Statistical Analysis

Statistical analysis is the process of applying statistical techniques on web log file to describe sessions, and user navigation such as viewing the time and length of a navigational path [Srivastava *et al.*, 2000]. Statistical prediction can also be used to predict when some page or document would be accessed from now [Dhyani *et al.*, 2002]. The work in [Borges and Levene, 1999] makes use of the *N-grammer* model which assumes that when a user is browsing a given page, the last *N* pages browsed affect the probability of the next page to be visited.

3.2 Clustering

Clustering is the process of partitioning a given population of events or items into sets of similar elements [Han and Kamber, 2001]. In web usage mining there are two main interesting clusters to be discovered: usage clusters, and pages clusters [Srivastava *et al.*, 2000]. The authors in [Su *et al.*, 2002] present an approach to cluster web pages to have a high quality clusters of web pages and use that clusters to produce index pages, where index pages are web pages that have direct links to pages that may be of interest of some group of website navigators. In [Koutri and Daskalaki, 2003] clustering techniques are applied to web log file to discover those subsets of web pages that need to be connected, and to improve the already connected pages. [Olga. *et al.*, August 1999] uses the *Competitive Agglomeration Clustering Algorithm* to cluster the sessions extracted from web log server into typical session profiles of users. The authors in [Velásquez *et al.*, 2003] use a clustering algorithm which identifies groups of similar sessions, allowing the analysis of visitor behavior.

3.3 Classification

Classification is dividing an existing set of events or transactions into another predefined sets or classes based on some characteristics. In web usage mining, classification is used to group users into a predefined groups with respect to their navigation patterns in order to develop profiles of users belonging to a particular class or category [Srivastava *et al.*, 2000]. [Ester *et al.*, 2002] introduces several approaches for web page classification. The authors in [Fu *et al.*, 2001] propose an approach to reorganize a website based on user access patterns and the classification of web pages into two categories: index pages, and content pages.

3.4 Association Rule Mining

Association rule mining is the discovery of attribute values that occur frequently together in a given set of data [Han and Kamber, 2001]. Association rules mining techniques are used in web usage mining to find pages that are often viewed together, or to show which pages tend to be visited within the same user session [Baron and Spiliopoulou, 2003]. The work introduced in [Xue *et al.*, 2002] proposes a re-ranking method with the help of website taxonomy to mine for generalized association rules and abstract access patterns of different levels to improve the performance of site search. The authors in [Yang *et al.*, 2002] propose an approach for predicting web log accesses based on association rule mining. Association rule mining facilitates the identification of related pages or navigation patterns which can be used in web personalization [Mobasher *et al.*, 2001][Mobasher *et al.*, 2000].

3.5 Sequential Pattern Mining

In sequential pattern mining a sequence of actions or events is determined with respect to time or other sequences [Velásquez *et al.*, 2003]. In web usage mining, sequential pattern mining could be used to predict user's future visit behaviors. Some web usage mining and analysis tools use sequential pattern mining to extract interesting patterns such as *SpeedTracer* [Wu *et al.*, 1998], and *WEBMINER* [Cooley *et al.*, 1997]. The authors in[Buchner *et al.*, 1999] suggest using adaptive websites to attract customers using sequential patterns to display special offers dynamically to them.

4 Data Preprocessing

Before data mining techniques are applied to web log file data, several preprocessing steps should be done in order to make web log file data ready to be mined. Web log file contains data about requested URL, time and date of request, method used, etc. The main data preprocessing tasks are data cleaning and filtering, path completion, session identification, and session formatting.

4.1 Data Cleaning

Data cleaning is the first preprocessing task. It involves the removal or elimination of irrelevant items that are not important for any type of web log analysis. Elimination of irrelevant items can be accomplished by checking the suffix of the URL name to filter out requests for graphics, sound, and video hits in order to concentrate on data representing actual page hits [Cooley *et al.*, 1997][Zaiane *et al.*, 1998]. For example, all log entries with filename suffixes such as gif, jpeg, and jpg can be removed. Another cleaning process is removing log entries generated by web agents like web spiders, indexers, or link checkers [Zaiane *et al.*, 1998]. Filtering out failed server requests, or transforming server error code is also done. Merging logs from multiple servers and parsing the log into data fields is also considered a data cleaning step [Cooley, 2003].

4.2 Path Completion

Path completion preprocessing task fills in page references that are missing due to local browsing caching such as using the back button available in the browser to go back to previously visited page [Cooley *et al.*, 1999].

4.3 User Identification

Identifying unique users is a complex step due to the existence of local caches, corporate firewalls, and proxy servers [Cooley *et al.*, 1997]. If the agent log shows a change in browser software, or operating system, a reasonable assumption to make is that each different IP address in the log file represent a different user [Pierrakos *et al.*, 2003]. If a page is requested that is not directly reachable by a hyperlink from any of the pages visited by the user, a heuristic assume that there is another user with the same IP address. Another assumption can be made is that consecutive accesses from the same host during a certain time interval come from the same user [Eirinaki and Vazirgiannis, 2003]. In some cases it is difficult to identify users, for example, when two users use the same machine and the same browser with the same IP address and look at the same set of pages [Cooley *et al.*, 1999].

4.4 Session Identification

A user session is defined as *the set of pages visited by the same user within the duration of one particular visit to a website* [Pierrakos *et al.*, 2003]. Session identification is dividing the page accesses of each user into individual sessions. One approach to identify user sessions, is by using a timeout threshold that is if the time between pages requests exceeds a certain limit (e.g. 30 minutes), then the user is starting a new session [Cooley *et al.*, 1999][Catledge and Pitkow, 1995]. Another approach assumes that consecutive accesses within the same time period belong to the same session [Eirinaki and Vazirgiannis, 2003].

4.5 Session Formatting

A final preprocessing step could be formatting the sessions or transactions for the type of the data mining technique, or algorithm to be applied [Cooley *et al.*, 1999]. The WEBMINER in [Cooley *et al.*, 1997] formats the cleaned web server log data in order to apply either association rule mining, or sequential pattern mining.

5 Web Usage Mining for Adaptive Websites

Adaptive websites are *websites that semi-automatically improve their organization and presentation by learning from user access patterns* [Perkowitz and Etzioni, 1998]. A site ability to adapt should be enhanced with information about its content, structure, and organization. For example, to add a link to a list of links ordered alphabetically, the link should be added at a specific point in the list. In the following subsections, we categorize different approaches of adaptive websites, even though it is difficult to make borders between different adaptation approaches for example, improving website links yields consequently to improve the structure of the website.

5.1 Improving Site Usability and Organization

Improving site usability can be achieved through making changes to the organization of the pages and links of the website. The work in [Fu *et al.*, 2001] aims to build an adaptive website that will reorganize its pages so that its users can find the information they want with minimum effort, where *effort* is defined in [Perkowitz and Etzioni, 1998] as *a function of the number of links traversed and the difficulty of finding that links in website pages*. Reorganization process is done by firstly extracting access patterns from web server's log file. Secondly, the web pages

in the web sever are classified into index pages and content pages based on the characteristics and access statistics of the pages. Finally, the whole website is analyzed and a reorganization of the website is presented based on access information and page classification. The authors in [Srikant and Yang, 2001] propose an algorithm to automatically find pages in a website whose location is different from where visitors expect to find them. The expected locations are then presented to the website administrator to add a navigation link from the expected location to the target page. The authors also present another algorithm to select the set of navigation links to optimize the benefit to the website or the visitor. The authors in [Spiliopoulou and Pohle, 2001] present a model to improve the success of the web site with the help of data mining techniques. To evaluate the efficiency values of a site pages, the authors analyze the navigational behavior of the site visitors with web usage mining. The analyst may decide to perform navigation pattern discovery over the entire log or to split it into customer log, or non-customer log and performs a comparative analysis of the two. Then makes decisions depending on the discovered results. The authors in [Mikroyannidis and Theodoulidis, 2005] introduced a framework that enables adaptation of the web topology and ontology to the needs and interests of web users. The proposed adaptation process exploits the access data of the users, together with the semantic aspect of the web, in order to facilitate web browsing.

5.2 Adaptive Content

Changing the content of a website can make the website better serve the requirements of a specific user. Content may be added, removed, or rearranged [Kilfoil *et al.*, 2003]. This includes additional explanations or details which may be added or removed depending on user's background and interests in some topic, or changing the website presentation language based on the user language preference.

5.3 Adaptive Link

Making changes to the links of the website can facilitate user's navigation of the website and minimize the time required to reach the target page. There are several techniques for adaptive link such as direct guidance, link sorting, link hiding, disabling, or highlighting. Direct guidance technique provides the user with a link to the page which is predicted to be the best next step for the user [Brusilovsky, 1997]. The AVANTI project [Fink *et al.*, 1996] tries to predict user's goals and presents links leading directly to pages it thinks a user will want to see. The work in [Wexelblat., 1996] proposes an approach to suggest a path to unexperienced users if many users follow the same path in their search for information. Link sorting is done by selecting the most relevant pages based on the users interests or goals then sorting them based on their relevance and presenting them in an ordered list of links [Kilfoil *et al.*, 2003][Brusilovsky, 1997]. Hiding or disabling the links that are not relevant to the user interests and goals makes the user less confused and speeds up user's navigation [Kilfoil *et al.*, 2003][Brusilovsky, 1996]. Link highlighting can also facilitate user's navigation [Brusilovsky, 1997][Brusilovsky, 1996].

5.4 Adaptive Web Structure

Adding or removing new pages is a final decision of the website administrator. Depending on the extracted usage patterns, several changes may be done on website structure. The authors in [Perkowitz and Etzioni, 1998] investigate the creation of index pages, which are pages that contain a direct link to pages that cover a particular topic, to facilitate the user's navigation of the website. The *PageGather* cluster mining algorithm is introduced. It takes web server logs as input and finds collections (clusters) of pages that tend to co-occur in visits, and outputs the contents of candidate index pages for each cluster found. A further development to [Perkowitz and Etzioni, 1998] is found in [Perkowitz and Etzioni, 2000] by presenting the *IndexFinder* a conceptual clustering mining algorithm in which all discovered clusters have intuitive descriptions that can be expressed to human users to solve the problem that *PageGather* gives no guarantee that all objects in the discovered cluster are about the same topic. To measure the use of a set of pages [Kilfoil *et al.*, 2003] statistics about commonly viewed pages, and subsets of pages is generated. The administrator can get an idea how the structure of the web should be, and whether there are some pages need to be removed, added, or their position need to be changed, without destroying the overall structure of the website.

5.5 Adaptive E-Commerce

Web usage mining has a great effect on e-commerce. It can be used to study customer behavior in the web, and use the extracted knowledge to facilitate navigation and services introduced to the customer, and suggest some particular products to the customer based on his interests. In [Berendt and Spiliopoulou, 2000] comparisons of navigation patterns between customers and non-customers lead to rules that specify how the website should be improved. The work in [Buchner *et al.*, 1999] suggests using adaptive websites to attract customers using sequential patterns to display special offers dynamically to them, and to keep the online shopper as loyal as possible. An example of e-commerce site that uses personalization is amazon.com, in which recommendations are presented to different customers depending on the customer profile [Au, 2002].

In order to make websites more effective to website users, they should reflect their interests, knowledge, needs, and goals. This can be done through personalization which is the subject of the next section.

6 Web Usage Mining for Personalized Websites

Web personalization is the process of customizing websites to the needs of specific users taking advantage from the patterns discovered from mining web usage data and other information such as web structure, web content, and user profile data [Eirinaki and Vazirgiannis, 2003]. Web personalization begins with the collection of web data. In this stage usage data are collected from different sources such as web server side data, client side data, and proxy servers. In general, personalization techniques are divided into offline, and online techniques. Offline personalization is based on simple user profiling and manual decision rule systems. Web usage mining is an online personalization data source. By evaluating site behavior and usage, a view about the website user is gained which yields to a more effective personalization strategies. User profiles are an important source of data for data personalization. User profiles contain user preferences, characteristics, interests knowledge, skills, activities, and behavioral patterns [Koutri *et al.*, 2005]. Such information is obtained

either explicitly using online registration forms and questionnaires resulting in static user profiles, or implicitly by recording the navigational behavior and/or the preferences of each user resulting in dynamic user profiles [Eirinaki and Vazirgiannis, 2003].
There are different ways to analyze the collected data. Content based filtering methods select content items that have a high degree of similarity to the user's profile [Vassiliou *et al.*, 2002]. An alternative to content based filtering is the collaborative filtering techniques which allow users to take advantage of other users behavioral activities based on a measure of similarity between them [Vassiliou *et al.*, 2002][Kim *et al.*, 2004]. Rule based filtering allows website administrators/marketers to specify business rules based on user demographics. The rules are used to affect the content introduced to a particular user.
Pattern discovery is the next step of the personalization process. In this step, different data mining techniques, such as clustering, classification, association rule mining, and sequential pattern analysis, are used to discover interesting patterns from web usage data.
Clustering is used to group users with common browsing behavior. The authors in [Shahabi *et al.*, 1997] implement a *Profiler* system which captures client's selected links, page order, page viewing time, and cache references. That information are used to cluster users with similar interests. The work in [Mobasher *et al.*, 2000] proposes a recommendation engine which considers the association rules between different web pages, and the derivation of URL clusters based on two types of clustering techniques in conjunction with the active user session. The recommendations are then added to the last requested page as a set of links before the page is sent to the client browser.
Association rules or sequential pattern discovery methods facilitate the identification of related pages or navigation patterns which can be used subsequently to recommend new web pages to the visitors of a website. The work in [Mobasher *et al.*, 2001] provides a framework for web personalization based on association rule mining from clickstream data. [Ishikawa *et al.*, 2002] introduces the *System L-R* recommendation system which constructs user models by classifying the web access and recommends relevant pages to the users based both on the user models and the web content.
[Pierrakos *et al.*, 2001] presents a web usage mining system *KOINOTITES* which uses web usage mining techniques to identify groups of users who have similar navigation behavior. The produced information can either be used by the administrator in order to improve the structure of the website or it can be fed directly to a personalization model, (e.g., collaborative filtering). The work in [Albanese *et al.*, 2004] proposes a web mining strategy for web personalization based on a novel pattern recognition strategy which analysis and classifies users taking into account both user provided data and navigational behavior of the users. It presents the *Refferer Based Page Recommendation, RBPR,* that uses information about a visitor's browsing context (specially, the refferer URL provided by the HTTP) to suggest pages that might be relevant to the visitors underlying information need.
The authors in [Kushmerick *et al.*, 2000] introduce a different approach of personalization that requires no input or feedback from the user. The work in [Vassiliou *et al.*, 2002] suggests a set of steps that make the personalization process effective starting from data collection and management efforts, to measuring and evaluating the success of personalization.

7 Selected Web Mining and Analysis Tools

In this section we present some well-known web usage mining and analysis tools such as WUM, SpeedTracer, Weblogminer, WebMiner, WebWatcher, and WebPersonalizer.

- ***Web Utilization Miner (WUM)*:** the system discovers the navigational patterns [Spiliopoulou and Faulstich, 1998]. A human expert specify the generic structural and statistical characteristics that makes a navigation pattern interesting to improve the organization of the web documents and adapt it better to the needs of users.
- ***SpeedTracer*:** SpeedTracer [Wu *et al.*, 1998] reconstructs user traversal paths to identify user sessions. It uses association rule mining and sequential patterns to present statistics about users, the most frequently visited pages, the distribution of user session durations, the number of visited pages, the most frequently traversed paths, and the most frequently visited groups of pages.
- ***WebLogMiner*:** WebLogMiner is a general web usage mining tool [Zaiane *et al.*, 1998]. It consists of four steps. The first step is filtering web log file and creating a relational database for the filtered information, containing different attributes such as user, resource, day, etc. In the second step, a data cube is constructed using the available dimensions. Then, online analytical processing OLAP techniques are used on the web log data cube. Finally, data mining techniques such as data characterization and comparison, statistical analysis, classification, and time series analysis are put to use with the data cube to predict, classify, and discover interesting correlations.
- ***WEBMINER*:** WEBMINER [Cooley *et al.*, 1997] uses sequential pattern mining, and association rule mining techniques. Before any knowledge discovery technique takes place, the web server log data is cleaned. The resulting data is then formatted according to the data mining technique need to be used. Instead of mining for all patterns, a Query Mechanism is used to limit the search to relevant and useful patterns. The resulting patterns are used for web restructuring, and personalization.
- ***WebWatcher*:** WebWatcher is a personalization web mining tool [Joachims *et al.*, 1997]. It is a tour guide agent that provides navigational hints to the user while browsing the website such as highlighting interesting links, based on user interests, and the content of the web pages. The system learns from the earlier tours to improve the recommendation giving skills.
- ***WebPersonalizer*:** after web log data are preprocessed data mining techniques such as association rule mining, sequential pattern discovery, clustering, and classification are applied to discover interesting patterns. The results are then used to create aggregated usage profiles in order to create decision rules. After matching each user activity and those usage profiles, a list of recommended links are provided to the user [Mobasher, 2001].

8 Summary and Research Direction

In this paper, we introduced an overview about various research issues in web usage mining and its application in the field of adaptive, and personalized web sites.
In this paper, different approaches for reorganizing and improving the design of the websites based on user navigational patterns, and user profiles have been discussed. Most of that approaches are semi-automatic i.e., they need user, or website administrator interaction in order to complete the adaptation process, such as questionnaires filled by user, or when some interesting navigational patterns are discovered, the website administrator use that patterns to make decisions to adapt the website. Different approaches try to make the adaptation procedure as automatic as possible. Other approaches try to find new measures that reflect further characteristics of website usage, and improve pattern analysis by improving visualization tools to make it easier for the analyst to understand the extracted patterns. A new approach to enhance web personalization is by making some psychological studies on user profiles which yields to better personalized websites.
Another approach for website adaptation is making simultaneous adaptation of multiple websites. That websites have something common with each other or belong to the same category of websites. Another similar approach is dealing with the website as a collection of gates. Each gate represent a general subject (i.e. sport, medicine, computer, etc.), and within every gate there are some recommended links for the web navigator. Those recommendations include links to related websites that may be of interest to the web navigator. The recommendations are done depending on the log history of previous users.

References

[Albanese *et al.*, 2004] Massimiliano Albanese, Antonio Picariello, Carlo Sansone, and Lucio Sansone. A Web Personalization System Based on Web Usage Mining Techniques. In *WWW Alt. 2004*, pages 288–289. ACM Press, 2004.

[Au, 2002] SaiMing Au. A Study of Application of Web Mining for E-Commerce: Tools and Methodology. *International Journal of The Computer, The Internet and Management*, 10(3):1–14, 2002.

[Baron and Spiliopoulou, 2003] Steffan Baron and Myra Spiliopoulou. Monitoring the Evolution of Web Usage Patterns. In *EWMF*, pages 181–200, 2003.

[Berendt and Spiliopoulou, 2000] Bettina Berendt and Myra Spiliopoulou. Analysis of Navigational Behavior in Web Sites Integrating Multiple Information System. *VLDB Journal*, 9:56–75, 2000.

[Borges and Levene, 1999] Jose Borges and Mark Levene. Data Mining of User Navigation Patterns. In *WEBKDD*, pages 92–111, 1999.

[Brusilovsky, 1996] Peter Brusilovsky. Methods and Techniques of Adaptive Hypermedia. *User Modeling and User-Adapted Interaction*, 6:87–129, 1996.

[Brusilovsky, 1997] Peter Brusilovsky. Efficient Techniques for Adaptive Hypermedia. In *Intelligent Hypertext*, pages 12–30, 1997.

[Buchner *et al.*, 1999] Alex Buchner, Maurice D. Mulvenna, Sarab S. Anand, and John G. Hughes. An Internet-enabled Knowledge Discovery Process. Technical report, MINEit Software Ltd., 1999.

[Cai *et al.*, 2004] Ding Cai, Xiaofei He, Ji-Rong Wen, and Wei-Ying Ma. Block-Level Link Analysis. In *Proc. of SIGIR 2004*, pages 440–447. ACM Press, 2004.

[Catledge and Pitkow, 1995] Lara D. Catledge and James E. Pitkow. Characterizing Browsing Strategies in the World Wide Web. *Computer Networks and ISDN Systems*, 27(6):1065–1073, 1995.

[Cohen, 2003] William W. Cohen. Learning and Discovering Structure in Web Pages. *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering*, 26(1):3–10, 2003.

[Cooley *et al.*, 1997] R. Cooley, J. Srivastava, and B. Mobasher. Web Mining: Information and Pattern Discovery on the World Wide Web. In *Proc. of ICTAI 1997*, page 558, Nov. 1997.

[Cooley *et al.*, 1999] Robert Cooley, Bamshad Mobasher, and Jaideep Srivastava. Data Preparation for Mining World Wide Web Browsing Patterns. *Knowledge and Information Systems*, 1(1):5–32, 1999.

[Cooley, 2003] Robert Cooley. The Use of Web Structure and Content to Identify Subjectively Interesting Web Usage Patterns. *ACM Trans. Inter. Tech.*, 3(2):93–116, 2003.

[Davison, 2002] Brian D. Davison. Predicting Web Actions from HTML Content. In *Proc. of the The 13th ACM Conf. on Hypertext and Hypermedia*, pages 159–168, College Park, MD, Jun 2002.

[Dhyani *et al.*, 2002] D. Dhyani, W. Keong, and N. Bhowmick. A Survey of Web Metrics. In *ACM Computing Surveys*, volume 34, pages 469–503, 2002.

[Eirinaki and Vazirgiannis, 2003] Magdalini Eirinaki and Michalis Vazirgiannis. Web Mining for Web Personalization. *ACM Transactions for Internet Technology*, 3(1):1–27, 2003.

[Ester *et al.*, 2002] Martin Ester, Hans-Peter Kriegel, and Matthias Schubert. Web Site Mining a New Way To Spot Competitors, Customers and Suppliers in The World Wide Web. In *Proc. 8th ACM SIGKDD KDD 2002*, pages 249–258, Edmonton, CA, 2002. ACM Press.

[Fink *et al.*, 1996] J. Fink, A. Kobsa, and A. Nill. User-oriented Adaptivity and Adaptability in the Avanti Project. In *In Proc. of the Designing for the Web Conf: Empirical Studies, Microsoft Campus Redmond, USA*, 1996.

[Fu *et al.*, 2001] Y. Fu, M. Creado, and M. Shih. Adaptive Web Sites by Web Usage Mining. In *Int. Conf. on Internet Computing (IC 2001)*, Las Vegas NA, 2001.

[Gedov *et al.*, 2004] Vassil Gedov, Carsten Stolz, Ralph Neuneier, Michal Skubacz, and Dietmar Seipel. Matching Web Site Structure and Content. In *Proc. WWW 2004*, pages 286–287. ACM Press, 2004.

[Han and Kamber, 2001] Jiawei Han and Micheline Kamber. *Data Mining Concepts and Techniques*. Morgan Kaufmann Publishers, San Francisco, 2001.

[Ishikawa *et al.*, 2002] Hiroshi Ishikawa, Manabu Ohta, Shohei Yokoyama, Junya Nakayama, and Kaoru Katayama. On the Effectiveness of Web Usage Mining for Page Recommendation and Restructuring. In *Web,*

Web-Services, and Database Systems, pages 253–267. Springer-Verlag, 2002.

[Joachims *et al.*, 1997] Thorsten Joachims, Dayne Freitag, and Tom M. Mitchell. WebWatcher: A Tour Guide for the World Wide Web. In *IJCAI (1)*, pages 770–777, 1997.

[Kilfoil *et al.*, 2003] M. Kilfoil, A. Ghorbani, W. Xing, Z. Lei, J. Lu, J. Zhang, and X. Xu. Toward an Adaptive Web: The State of the Art and Science. In *Proc. of CNSR 2003*, pages 119–130, Moncton, Canada, 2003.

[Kim *et al.*, 2004] Dong-Ho Kim, Vijayalakshmi Atluri, Michael Bieber, Nabil Adam, and Yelena Yesha. A Clickstream-based Collaborative Filtering Personalization Model: Towards a Better Performance. In *WIDM 2004*, pages 88–95. ACM Press, 2004.

[Kosala and Blockeel, 2000] Raymond Kosala and Hendrik Blockeel. Web mining research: A survey. *ACM SIGKDD Explorations*, 2(1):1–15, 2000.

[Koutri and Daskalaki, 2003] Martha Koutri and Sophia Daskalaki. Improving Web Site Usability Through a Clustering Approach. In *10th International Conference on Human-Computer Interaction HCI, Crete, Greece*, pages 11–19, 2003.

[Koutri *et al.*, 2005] M. Koutri, N. Avouris, and S. Daskalaki. A Survey on Web Usage Mining Techniques for Web-based Adaptive Hypermedia Systems. In *Adaptable and Adaptive Hypermedia Systems, IRM Press*, pages 125–149, 2005.

[Kushmerick *et al.*, 2000] Nicholas Kushmerick, James McKee, and Fergus Toolan. Towards Zero-Input Personalization: Referrer-Based Page Prediction. *Lecture Notes in Computer Science*, 1892:133–143, 2000.

[Lin and Ho, 2002] Shian-Hua Lin and Jan-Ming Ho. Discovering Informative Content Blocks from Web Documents. In *Proc. ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, pages 588–593. ACM Press, 2002.

[Liu *et al.*, 2001] Bing Liu, Yiming Ma, and Philip S. Yu. Discovering Unexpected Information from Your Competitors' Web Sites. In *Proc. 7th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, pages 144–153. ACM Press, 2001.

[Liu *et al.*, 2003] Bing Liu, Robert Grossman, and Yanhong Zhai. Mining Data Records in Web Pages. In *Proc. 9th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, pages 601–606. ACM Press, 2003.

[Mikroyannidis and Theodoulidis, 2005] Alexander Mikroyannidis and Babis Theodoulidis. Web Usage Driven Adaptation of the Semantic Web. In *End User Aspects of the Semantic Web Workshop, 2nd European Semantic Web Conference (ESWC 2005)*, Heraklion, Greece, 2005.

[Mobasher *et al.*, 2000] Bamshad Mobasher, Robert Cooley, and Jaideep Srivastava. Automatic Personalization Based on Web Usage Mining. *Commun. ACM*, 43(8):142–151, 2000.

[Mobasher *et al.*, 2001] Bamshad Mobasher, Honghua Dai, Tao Luo, and Miki Nakagawa. Effective Personalization Based on Association Rule Discovery from Web Usage Data. In *WIDM 2001*, pages 9–15, NY, USA, 2001. ACM Press.

[Mobasher, 2001] Bamshad Mobasher. WebPersonalizer: A Server-Side Recommender System Based on Web Usage Mining. Technical Report TR01-010, School of Computer Science, Telecommunications and Information Systems, DePaul University, Chicago, IL, USA, 2001.

[Morinaga *et al.*, 2002] Satoshi Morinaga, Kenji Yamanishi, Kenji Tateishi, and Toshikazu Fukushima. Mining Product Reputations on the Web. In *Proc. of the 8th ACM SIGKDD int. conf. on Knowledge discovery and data mining*, pages 341–349. ACM Press, 2002.

[Olga. *et al.*, August 1999] Nasraoui Olga., H. Frigui, A. Joshi, and R. Krishnapuram. Mining Web Access Logs Using Relational Competitive Fuzzy Clustering. In *Proc. of the 8th Int. Fuzzy Systems Association Congress*, Hsinchu, Taiwan, August 1999.

[Perkowitz and Etzioni, 1998] Mike Perkowitz and Oren Etzioni. Adaptive Web Sites: Automatically Synthesizing Web Pages. In *Proc. 15th National Conference on Artificial Intelligence and 10th Innovative Applications of Artificial Intelligence Conference (AAAI 98 / IAAI 98)*, pages 727–732. AAAI Press/The MIT Press, 1998.

[Perkowitz and Etzioni, 2000] Mike Perkowitz and Oren Etzioni. Adaptive Web Sites. *Communications of the ACM*, 43(8):152–158, 2000.

[Pierrakos *et al.*, 2001] Dimitrios Pierrakos, Geogios Paliouras, Christos Papatheodouro, and Constantine D. Spyropoulos. KOINOTITES: A Web Usage Mining Tool for Personalization. In *Proc. of the Panhellenic Conf. on Human Computer Interaction*, 2001.

[Pierrakos *et al.*, 2003] Dimitrios Pierrakos, Georgios Paliouras, Christos Papatheodorou, and Constantine D. Spyropoulos. Web Usage Mining as a Tool for Personalization: A Survey. *User Modeling and User-Adapted Interaction*, 13(4):311–372, 2003.

[Shahabi *et al.*, 1997] C. Shahabi, A. M. Zarkesh, J. Adibi, and V. Shah. Knowledge Discovery from Users Web Page Navigation. In *Proc. of the 7th Int. Workshop on Research Issues in Data Engineering RIDE 1997*, page 20, Washington, DC, USA, 1997. IEEE Computer Society.

[Spiliopoulou and Faulstich, 1998] Myra Spiliopoulou and Lukas C. Faulstich. WUM: a Web Utilization Miner. In *Workshop on the Web and Data Bases (WebDB 1998)*, pages 109–115, 1998.

[Spiliopoulou and Pohle, 2001] Myra Spiliopoulou and Carsten Pohle. Data Mining for Measuring and Improving the Success of Web Sites. *Data Min. Knowl. Discov.*, 5:85–114, 2001.

[Srikant and Yang, 2001] Ramakrishnan Srikant and Yinghui Yang. Mining Web Logs to Improve Web Site Organization. In *Proc. WWW 2001*, pages 430–437, Hong Kong, 2001. ACM Press.

[Srivastava *et al.*, 2000] Jaideep Srivastava, Robert Cooley, Mukund Deshpande, and Pang-Ning Tan. Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data. *ACM SIGKDD Explorations*, 1(2):12–23, 2000.

[Su *et al.*, 2002] Zhong Su, Qiang Yang, Hong Jiang Zhang, Xiaowei Xu, Yu-Hen Hu, and Shaoping Ma.

Correlation-Based Web Document Clustering for Adaptive Web Interface Design. *Knowledge and Information Systems*, 4(2):151–167, 2002.

[Vassiliou *et al.*, 2002] Charalampos Vassiliou, Dimitris Stamoulis, and D. Martakos. The Process of Personalizing Web Content: Techniques, Workflow and Evaluation. In *Proc. of the SSGRR 2002w*, 2002.

[Velásquez *et al.*, 2003] Juan Velásquez, Hiroshi Yasuda, and Terumasa Aoki. Combining the Web Content and Usage Mining to Understand the Visitor Behavior in a Web Site. In *Proc. ICDM 2003*, pages 669–672. IEEE Computer Society Press, 2003.

[Wexelblat., 1996] Alan Wexelblat. An Environment for Aiding Information-browsing Tasks. In *In Proc. of AAAI Spring Symposium on Acquisition, Learning and Demonstration: Automating Tasks for Users*, Birmingham, UK, 1996. AAAI Press.

[Wu *et al.*, 1998] K.-L. Wu, P. S. Yu, and A. Ballman. SpeedTracer: A Web Usage Mining and Analysis Tool. *IBM Systems Journal*, 37(1), 1998.

[Xue *et al.*, 2002] Gui-Rong Xue, Hua-Jun Zeng, Zheng Chen, Wei-Ying Ma, and Chao-Jun Lu. Log Mining to Improve the Performance of Site Search. In *Proc. WISE 2002 Workshops*, pages 238–245, Singapore, 2002. IEEE Computer Society Press.

[Yang *et al.*, 2002] H. Yang, S. Parthasarathy, and S. Reddy. On the Use of Constrained Associations for Web Log Mining. In *WEBKDD Workshop: Web Mining for Usage Patterns and User Profiles*, pages 100–118, 2002.

[Zaiane *et al.*, 1998] Osmar R. Zaiane, Man Xin, and Jiawei Han. Discovering Web Access Patterns and Trends by Applying OLAP and Data Mining Technology on Web Logs. In *Proc. of ADL 1998*, pages 1–9, Washington, DC, USA, 1998. IEEE Computer Society.